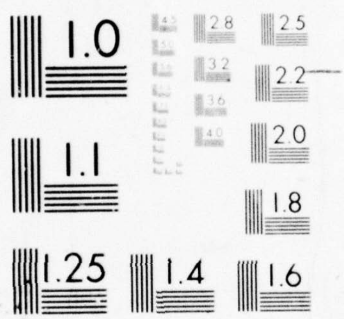
AD-A031 971 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
TESTING FOR EQUIPROBABILITY USING A MULTI-NOMIAL OR POISSON SAM--ETC(U)
JUL 76 L HOLST DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1659 NL

1 OF 1
AD
A031971

END
DATE
FILMED
1 -77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

11

FG

MRC Technical Summary Report #1659

TESTING FOR EQUIPROBABILITY USING A MULTINOMIAL OR POISSON SAMPLE

Lars Holst

See 1473

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

July 1976

(Received July 6, 1976)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

TESTING FOR EQUIPROBABILITY USING A MULTINOMIAL OR POISSON SAMPLE

Lars Holst

Technical Summary Report # 1659
July 1976

ABSTRACT

The well-known chi-square test is discussed for testing equiprobability when the expected number of observations per cell is not large. The results are used to give a justification for the Poisson Index of Dispersion test of fit for the Poisson distribution. A chi-square type statistic is studied for testing equiprobability and Poisson-fit when the frequencies associated with zero are missing and thus the sample size is unknown. Some applications are discussed in detail where the samples are incomplete.

AMS (MOS) Subject Classifications: Primary 62F05, Secondary 62E20

Key Words: Chi-square test, Poisson Index of Dispersion, Capture-recapture theory, Incomplete samples, Goodness-of-fit, Poisson distribution, Multinomial distribution

Work Unit Number 4 (Combinatorics, Probability, Statistics)

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

TESTING FOR EQUIPROBABILITY USING A MULTINOMIAL OR POISSON SAMPLE

Lars Holst

1. Introduction

Consider n independent replicates of an experiment with N different outcomes. An old statistical problem is to test for the same probability, $1/N$, for the outcomes. In Section 2 we will make some remarks about the well-known chi-square statistic's distribution when the expected number of observation per outcome, n/N, is not large. In Section 3 we will discuss testing fit for the Poisson distribution using the Poisson Index of Dispersion, i.e. comparing the sample mean and the sample variance. A rigorous proof of its asymptotic distribution is given using the results of Section 2. In some applications complete samples are not obtained. In Section 4 we will consider the multinomial and Poisson case when the frequencies associated with zero are missing and therefore N is unknown. A limit theorem is derived for a statistic of chi-square type, which can be used for testing equiprobability for the multinomial or fit for the Poisson distribution. In the last section some applications are discussed in detail where incomplete samples are obtained.

Some words about notation. A multinomial distribution with n repetitions and N equiprobable outcomes is denoted by $\mathrm{Mult}(n, 1/N, \ldots, 1/N)$.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

$Po(\lambda)$ stands for the Poisson distribution with mean λ, $N(m, \sigma^2)$ normal with mean m and variance σ^2, and $\chi^2(f)$ the usual chi-square with f degrees of freedom. Convergence in distribution is denoted by $\mathcal{L}(\cdots) \to \cdots$. To formulate limit theorems properly we should use an extra index ν, but to facilitate notation we suppress it.

2. Chi-square with equal probabilities

Let $(\xi_1, \ldots, \xi_N)$ be $\mathrm{Mult}(n, 1/N, \ldots, 1/N)$. Consider Karl Pearson's well-known chi-square statistic:

$$X^2_{n,N} = N \sum_{k=1}^{N} \xi_k^2/n - n . \tag{2.1}$$

For fixed N the following holds:

$$\mathcal{L}(X^2_{n,N}) \to \chi^2(N-1), \quad n \to \infty . \tag{2.2}$$

A result due to R. A. Fisher, correcting Pearson's mistake for the degrees of freedom, see e.g. Rao [14] for a proof.

The chi-square test is sometimes used in such a way that it is more logical to consider asymptotic results when both $n, N \to \infty$ in such a manner that $n/N \to \lambda$, $0 < \lambda < \infty$. Using different methods this case has been investigated by Harris and Park [7], Holst [8, 9] and Morris [13], showing that

$$\mathcal{L}((X^2_{n,N} - (N-1))/(2(N-1))^{1/2}) \to N(0,1) . \tag{2.3}$$

Note the independence of λ. As we have

$$\mathcal{L}((\chi^2(f) - f)/(2f)^{1/2}) \to N(0,1), \quad f \to \infty, \tag{2.4}$$

it is reasonable to use the chi-square approximation even if n/N, the expected number of observations/class, is small. This theoretical argument is not a consequence of the usual chi-square theory as given in e.g. [14]. By comparing moments of $X^2_{n,N}$ with those of $\chi^2(N-1)$,

see Rao and Chakravarti [15], one can expect the chi-square approximation to be accurate, better than the normal indicated by the limit result above, even if n/N is quite small. This is also confirmed by different numerical investigations, see Good, Gover and Mitchell [4], Katti [11], and Zahn and Roberts [18].

3. The Poisson Index of Dispersion

Let the random variables $\eta_1, \ldots, \eta_N$ be i.i.d. $Po(\lambda)$. The Poisson Index of Dispersion is defined as

$$T_N = \sum_1^N (\eta_k - \bar{\eta})^2 / \bar{\eta} = N \sum_1^N \eta_k^2 / \sum_1^N \eta_k - \sum_1^N \eta_k . \qquad (3.1)$$

It is a well-known fact that

$$(\eta_1, \ldots, \eta_N) \,\Big|\, \sum_1^N \eta_k = n \sim \mathrm{Mult}(n, 1/N, \ldots, 1/N) . \qquad (3.2)$$

Thus, using the notation of the previous section,

$$\mathfrak{L}(T_N \,\Big|\, \sum_1^N \eta_k = n) = \mathfrak{L}(X^2_{n,N}) . \qquad (3.3)$$

We can use this to construct a test for the hypothesis $H_0 : \eta_1, \cdots, \eta_N$ i.i.d. $Po(\lambda)$; consider $\sum_1^N \eta_k = n$ as given, reject H_0 if $T_N > c$ = constant. This test is an old one, it was introduced by R. A. Fisher in the first edition of "Statistical Methods for Research Workers". It is discussed in many papers see e.g. Haight [5, p. 94], Kathirgamatamby [10], and Rao and Chakravarti [14]. Often it is claimed that from the usual chi-square theory it follows that $c = \chi^2_\alpha(N-1)$, the upper α-percentile of chi-square, gives a test with approximate level α in large samples. This argument is not correct as N is not regarded as fixed. But, by the strong law of large numbers, with probability one,

$$\sum_1^N \eta_k / N = n/N \to \lambda , \qquad (3.4)$$

when the number of observations $N \to \infty$. Therefore we actually have the situation discussed in Section 2. The results there indicate that $c = \chi_\alpha^2(N-1)$ is a good approximation for large N. The significance level of the unconditional test: reject H_0 if $T_N > \chi_\alpha^2(N-1)$, is given by

$$\sum_{n=0}^{\infty} P(T_N > \chi_\alpha^2(N-1) \mid \sum_1^N \eta_k = n) P(\sum_1^N \eta_k = n) \to \alpha , \tag{3.5}$$

when $N \to \infty$. Thus we have obtained a correct theoretical justification for R. A. Fisher's old test based on the Poisson Index of Dispersion.

4. Empty classes missing

Consider the same model as in Section 2 i.e. $(\xi_1, \ldots, \xi_N)$ is $\mathrm{Mult}(n, 1/N, \ldots, 1/N)$. Here we suppose that the ξ's equal to 0 are not observed, N is unknown and the ordering of the ξ's is imagined. We want to test the model. This type of problem is of practical importance as we will see from the examples in the next section.

The limit result of Section 2, when $n/N \to \lambda$, can be stated as

$$\mathcal{L}((N \sum_{1}^{N} \xi_k^2/n - n - N)/(2N)^{1/2}) \to N(0,1) \ . \tag{4.1}$$

We could contemplate using an estimate of N in this formula. Essentially the maximum likelihood estimator of n/N is given by

$$(1 - \exp(-(n/N)^*))/(n/N)^*) = \sum_{1}^{N} I(\xi_k \neq 0)/n \ , \tag{4.2}$$

see Darroch [3], Harris [6], Lewontin and Prout [12], Samuel [16], or Seber [17, p. 136]. Table A. 3 in [17] is useful for solving the equation. The following limit theorem holds.

Theorem. If

$$(\xi_1, \ldots, \xi_N) \sim \mathrm{Mult}(n, 1/N, \ldots, 1/N) \ , \tag{4.3}$$

$$\lambda_N = n/N \to \lambda, \ 0 < \lambda < \infty, \ n, N \to \infty \ , \tag{4.4}$$

$$(1 - e^{-\lambda^*})/\lambda^* = \sum_{1}^{N} I(\xi_k \neq 0)/n \ , \tag{4.5}$$

then, when $n, N \to \infty$,

$$\mathcal{L}(Z_N) \to N(0,1) \ , \tag{4.6}$$

where

$$Z_N = \left(\sum_1^N \xi_k^2/\lambda^* - n - n/\lambda^*\right)\left(\lambda^*/(n(2 - \lambda^{*2}/(e^{\lambda^*} - 1 - \lambda^*)))\right)^{1/2} . \tag{4.7}$$

Remark 1. Note that $\sum_1^N \xi_k^2$, λ^* and n are observable quantities even if N and the θ's are unknown.

Proof. Using the equation (4.5) and the lemma below (cf. Remark 3) and well-known convergence theorems, see e.g. Rao [14, section 6a.2], we obtain that the random variable

$$Z_{1N} = \left(\sum_1^N \xi_k^2/\lambda^* - n - n/\lambda^*\right)/N^{1/2} \tag{4.8}$$

has the same asymptotic distribution as

$$Z_{2N} = (\varepsilon_{1N} + \varepsilon_{2N} \cdot \lambda_N^2 e^{\lambda_N}/(e^{\lambda_N} - 1 - \lambda_N))/\lambda_N \tag{4.9}$$

where

$$\varepsilon_{1N} = \sum_1^N (\xi_k^2 - \lambda_N - \lambda_N^2)/N^{1/2} , \tag{4.10}$$

$$\varepsilon_{2N} = \sum_1^N (I(\xi_k \neq 0) - (1 - e^{-\lambda_N}))/N^{1/2} . \tag{4.11}$$

From the lemma we find

$$\mathcal{L}(Z_{2N}) \rightarrow N(0, 2 - \lambda^2/(e^\lambda - 1 - \lambda)) . \tag{4.12}$$

As $\lambda^* \rightarrow \lambda$ in probability it follows that

$$\lambda^*/(2 - \lambda^{*2}/(e^{\lambda^*} - 1 - \lambda^*)) \rightarrow \lambda/(2 - \lambda^2/(e^\lambda - 1 - \lambda)) , \tag{4.13}$$

in probability. Therefore

$$\mathfrak{L}(Z_N) = \mathfrak{L}(Z_{1N} \cdot (\lambda^*/(\lambda_N(2 - \lambda^{*2}/(e^{\lambda^*} - 1 - \lambda^*))))^{1/2}) \to N(0,1) , \qquad (4.14)$$

which proves the assertion. ∎

Lemma. Let s, t be real numbers. If $\lambda_N = n/N \to \lambda$, $0 < \lambda < \infty$, then

$$\mathfrak{L}((s \sum_1^N (\xi_k^2 - \lambda_N - \lambda_N^2) + t \sum_1^N (I(\xi_k \neq 0) - (1 - e^{-\lambda_N})))N^{-1/2})$$

$$\to N(0, 2s^2\lambda^2 - 2st\lambda^2 e^{-\lambda} + t^2 e^{-2\lambda}(e^{\lambda} - 1 - \lambda)) . \qquad (4.15)$$

Proof. The assertion is a special case of general theorems on multinomial sums, see Harris [7], Holst [8,9].

Remark 2. Taking $t = 0$, $s = 1$ gives the limit result of Section 2.

Remark 3. Taking $s = 0$, $t = 1$ gives

$$\mathfrak{L}(\sum_1^N (I(\xi_k \neq 0) - (1 - e^{-\lambda_N}))/N^{1/2}) \to N(0, e^{-2\lambda}(e^{\lambda} - 1 - \lambda)) , \qquad (4.16)$$

a limit theorem for the classical occupancy problem.

As in Section 3 we can study $\eta_1, \ldots, \eta_N$ i.i.d. $Po(\lambda)$, random variables. Suppose that only non-zero η's are observed and N is unknown. To construct a test for the Poisson model we can consider the conditional distribution

$$(\eta_1, \ldots, \eta_N) \mid \sum_1^N \eta_k = n \sim (\xi_1, \ldots, \xi_N) \sim \text{Mult}(n, 1/N, \ldots, 1/N) , \qquad (4.17)$$

and use the result above (by the law of large numbers $\sum_1^N \eta_k/N \to \lambda$).

5. Applications where zeros are missing

We will discuss in detail three examples from completely different areas of application to indicate that the problems considered in Section 4 are of practical interest.

Example 1 (Biology). In Craigh [1] a method of estimating the size of a butterfly population is considered, see also [17, p. 137]. The butterflies were caught one at a time and released after marking. A total number of 341 different butterflies were caught of which 258 occurred once, 72 twice and 11 three times in the sample. We can imagine that the butterflies are numbered $1, 2, \ldots, N$, and we want to estimate N. Let ξ_k denote the number of times butterfly k has been caught. A possible model is that $(\xi_1, \ldots, \xi_N) \sim \mathrm{Mult}(n, 1/N, \ldots, 1/N)$. For the data above we have

$$n = \sum_1^N \xi_k = 258\cdot 1 + 62\cdot 2 + 11\cdot 3 = 435 ,$$

$$\sum_1^N I(\xi_k \neq 0) = 258 + 62 + 11 = 341 , \tag{5.1}$$

$$\sum_1^N \xi_k^2 = 258\cdot 1 + 72\cdot 4 + 11\cdot 9 = 645 .$$

To test the model the theorem in Section 4 can be used in the following way. First

$$(1 - e^{-\lambda^*})/\lambda^* = 341/435 , \tag{5.2}$$

and from table A.3 in [17] we find $\lambda^* = 0.5084$. After some computing we obtain $Z_N = -1.32$ which can be compared by a $N(0,1)$-percentile. As a big value indicates departure from assumptions we have no reason to reject the model. It is also possible to estimate the expected number of butterflies caught 1, 2, 3, 4 times. One finds 261.7, 66.5, 11.3, 1.4, close to the observed values. In [1] and [17, p. 157] it is suggested that these numbers can be used in a chi-square test. To make a rigorous mathematical justification for this is not an easy matter. Besides there is always the arbitrariness about pooling classes. This problem does not occur for the test above. In [17, p. 14], it is claimed that the Poisson Index of Dispersion test is more sensitive than the usual chi-square. One may conjecture that this also holds true for the test above compared with chi-square. This matter needs further investigation.

Example 2 (Medicine). In an example given in Dahiya and Gross [2], referring to an epidemic of cholera in a village of India, the incomplete Poisson-model of Section 4 was used. With $f_x = \#\{k; \eta_k = x\}$ we have $f_1 = 32$, $f_2 = 16$, $f_3 = 6$, $f_4 = 1$ and $f_x = 0$ for $x \geq 5$. Computing as in Example 1 we find $\lambda^* = 0.9722$ and $Z_N = -0.51$ indicating good fit. The missing number f_0 is estimated to be $33 = \lfloor 33.46 \rfloor$. From Remark 3 in Section 4 we can easily construct confidence intervals for N and therefore for n_0. For a 95% confidence interval we just solve

$$(1 - e^{-\lambda})/\lambda = \sum_{1}^{N} I(\xi_k \neq 0)/n \pm 1.96 \cdot ((e^{\lambda^*} - 1 - \lambda^*)/e^{2\lambda^*} n\lambda^*)^{1/2} . \tag{5.2}$$

We obtain $\lambda_\ell = 0.742$, $\lambda_u = 1.240$ and the confidence limits for N are $N_u = 86/0.742 = 116$ and $N_\ell = 69$. The 95% confidence interval for n_0 is $(69-55, 116-55) = (14, 61)$, slightly different from that given in [2]. As the distribution of $\sum_1^N I(\xi_k \neq 0)$ is probably better approximated by the normal than N^*'s, the method used here may be more accurate than using normal approximation of N^* as in [2].

Example 3 (History and Mintage). In a project on South Asian Monetary History at the University of Wisconsin-Madison the "Qunduz Hoard" is studied. This is a hoard of silver coins of the Indo-Greek kings of ancient India (Circa 200-100 B.C.). It derives its name, the Qunduz hoard, from its find spot in Afghanistan and is especially significant as it is one of the largest intact samples of the coinage of the Indo-Greeks available for study, and has been published with excellent photographic plates enabling the researcher to detect die differences accurately. The coins were produced using a certain tool, called a die, which was worn out after a certain number of coins. This number is approximately known and is different for the two sides of the coin, the so called obverse and reverse sides. One can distinguish between coins produced by different dies. Knowing the number of dies used to produce a certain coin type, the total number of coins of this type produced by the mint could be estimated, giving a measure of the economic activity of the period. The problem is therefore to estimate the number of dies.

In a sample let ξ_k denote the number of coins produced by die k. If the coins in the "Qunduz Hoard" can be considered as a random sample from the population of coins used in those days, the dies were used about the same number of times (much larger than the sample sizes) and the classification of the coins is accurate, then $(\xi_1, \ldots, \xi_N) \sim \text{Mult}(n, 1/N, \ldots, 1/N)$, would be an adequate probabilistic model. Here N is the unknown number of dies, the zeros are not observed and the labeling is imagined. Using the method described above we can estimate N. The underlying assumptions implying equiprobability are crucial, so the model should be tested for fit. This can be done as in the other examples. The Poisson model would also be reasonable in this example. The same statistical analysis is used in either model.

As an illustration let us consider the following data from the "Qunduz Hoard". Using the notation of Example 2 we have for "obverse" sides of "Heliocles" coins: $f_1 = 102$, $f_2 = 26$, $f_3 = 8$, $f_4 = 2$, $f_5 = f_6 = f_7 = 1$, $f_x = 0$ for $x \geq 8$. We find $\lambda^* = 0.7907$ and using this $N^* = 258$. But $Z_N = 6.15$, a highly significant value indicating bad fit. Estimates of expected values are 92.5, 36.5, 7.6, 1.9, 0.3, 0.04, 0.004 showing that $f_6 = f_7 = 1$ are unlikely in a random sample. Adopting the standard rule for chi-square, that the expected number of observations in a class should be at least 5, we must pool classes 3-7 giving a chi-square of 4.17 with $3 - 1 - 1 = 1$ degree of freedom. As $\chi^2_{0.05}(1) = 3.84 < 4.17 < \chi^2_{0.01}(1) = 6.63$ there is significant deviation but just on the 5%-level;

cf. the conjecture in the end of Example 1. As the underlying model is very much in doubt one can expect the estimate of N to be biased. For the reverse side of the same coins we have: $f_1 = 156$, $f_2 = 19$, $f_3 = 2$, $f_4 = 1$, $f_x = 0$ for $x \geq 5$. From the same statistical analysis we get $\lambda^* = 0.2792$, $N^* = 731$, $Z_N = 1.57$ and expected values 154.4, 21.6, 2.0, 0.14. So the fit is adequate. A 95% confidence interval for N is (536, 1107). We can remark that the difference of fit between the "obverse" and "reverse" sides can be explained. A plausible explanation is, that the number of coins produced by the same die varies much more for the obverse side, because it was much more difficult to make an obverse die than a reverse.

I would like to thank Mr. John Deyell, Department of South Asian History, and Mr. Richard Bittman, Department of Mathematics and the Statistical Laboratory, University of Wisconsin-Madison, for providing me with this example and for discussions on it. The inspiration for this paper arose largely in connection with this application.

REFERENCES

[1] Craigh, C. C., "On the Utilization of Marked Specimens in Estimating Populations of Flying Insects," Biometrika, 40 (June 1953), 170-176.

[2] Dahiya, R. C. and Gross, A. J., "Estimating the Zero Class from a Truncated Poisson Sample," Journal of the American Statistical Association, 68 (September 1973), 731-733.

[3] Darroch, J. N., "The Multiple-Recapture Cencus. I. Estimation of a Closed Population," Biometrika, 45 (December 1958), 343-359.

[4] Good, I. J., Gover, T. N. and Mitchell, G. J., "Exact Distributions for χ^2 and for the Likelihood-Ratio Statistics for the Equiprobable Multinomial Distribution," Journal of the American Statistical Association, 65 (March 1970), 267-283.

[5] Haight, F. A., Handbook of the Poisson Distribution, New York: Wiley, 1967.

[6] Harris, B., "Statistical Inference in the Classical Occupancy Problem Unbiased Estimation of the Number of Classes," Journal of the American Statistical Association, 63 (September 1968), 837-847.

[7] Harris, B. and Park, C. J., "The Distribution of Linear Combinations of the Sample Occupancy Numbers," Indigationes Mathematicae 33, No. 2 (1971), 121-134.

[8] Holst, L., "Asymptotic Normality and Efficiency for Certain Goodness-of-Fit Tests," Biometrika 59 (April 1972), 137-145.

[9] Holst, L., "On Multinomial Sums," Technical Summary Report No. 1629, Mathematics Research Center, University of Wisconsin-Madison, 1976, submitted to the Annals of Statistics.

[10] Kathirgamatamby, N., "Note on the Poisson Index of Dispersion," Biometrika 40 (June 1953), 225-228.

[11] Katti, S. K., "Exact Distribution for the Chi-Square Test in the One Way Table," Communications in Statistics 2, No. 5 (1973), 435-447.

[12] Lewontin, R. C., and Prout, T., "Estimation of the Number of Different Classes in a Population," Biometrics 12 (June 1956), 211-223.

[13] Morris, C., "Central Limit Theorems for Multinomial Sums," Annals of Statistics 3 (January 1975), 165-188.

[14] Rao, C. R., Linear Statistical Inference and Its Applications, 2nd ed., New York: Wiley, 1973.

[15] Rao, C. R. and Chakravarti, I. M., "Some Small Sample Tests of Significance for a Poisson Distribution," Biometrics 12 (September 1956), 264-282.

[16] Samuel, E., "Comparison of Sequential Rules for Estimation of the Size of a Population," Biometrics 25 (September 1969), 517-527.

[17] Seber, G. A. F., The Estimation of Animal Abundance and Related Parameters, London: Griffin, 1973.

[18] Zahn, D. A. and Roberts, G. C., "Exact χ^2 Criterion Tables with Cell Expectations One: An Application to Coleman's Measure of Consensus," Journal of the American Statistical Association 66 (March 1971), 145-148.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
1659

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 TESTING FOR EQUIPROBABILITY USING A MULTINOMIAL OR POISSON SAMPLE.

5. TYPE OF REPORT & PERIOD COVERED
Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
9 Lars/Holst

8. CONTRACT OR GRANT NUMBER(s)
15 DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE
11 Jul[illegible]976

13. NUMBER OF PAGES
16

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
9 Technical summary rept.,

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
12 19p.

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
14 MRC-TSR-1659

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Chi-square test
Poisson Index of Dispersion
Capture-recapture theory
Incomplete samples
Goodness-of-fit
Poisson distribution
Multinomial distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
The well-known chi-square test is discussed for testing equiprobability when the expected number of observations per cell is not large. The results are used to give a justification for the Poisson Index of Dispersion test of fit for the Poisson distribution. A chi-square type statistic is studied for testing equiprobability and Poisson-fit when the frequencies associated with zero are missing and thus the sample size is unknown. Some applications are discussed in detail where the samples are incomplete.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) LB